

Frequency Band-Importance Functions for Auditory and Auditory-Visual Speech Recognition

*Ken W. Grant
Walter Reed Army Medical
Center
Washington, D.C. 20307-5001*

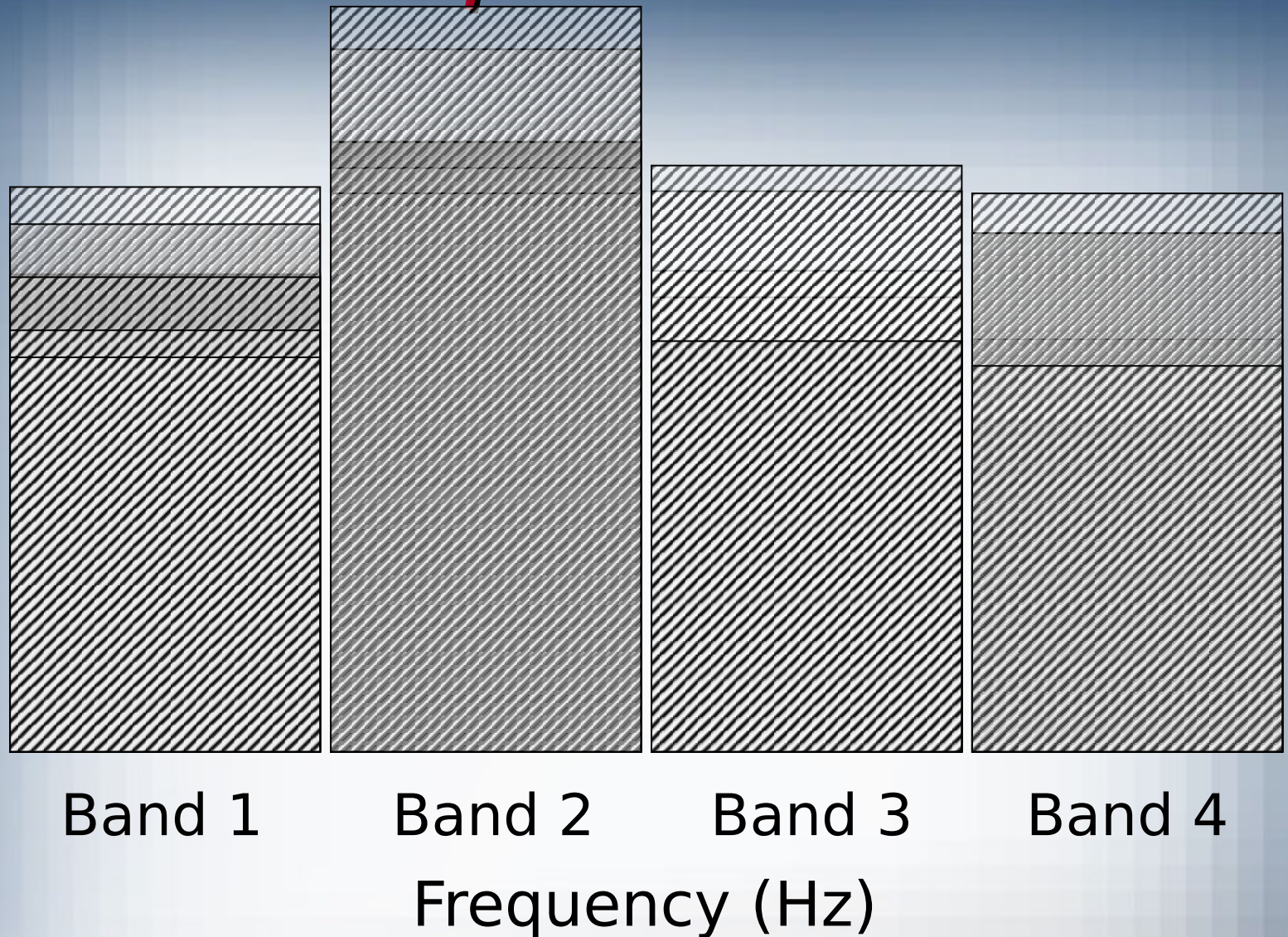
Background

- Speech recognition involves broadband listening.
- Information is not uniformly distributed across the frequency spectrum.
 - different cues (spectral and temporal) of different relative value reside at different frequencies.
 - in general, more importance is placed at mid-frequencies around 1500-3000 Hz.
 - probably related to place-of-articulation cues (F2/F3 transitions)

Background (continued)

- How can we determine the relative importance or “weights” that listeners place on various frequency regions?
- Doherty and Turner, 1996; Turner et al., 1998
 - correlational procedure (Lutfi, 1995; Richards and Zhu, 1994) applied to speech recognition.
 - partition speech into a number of spectral bands.
 - perturb each band so that amount of information in each band can be correlated with a listener’s performance.

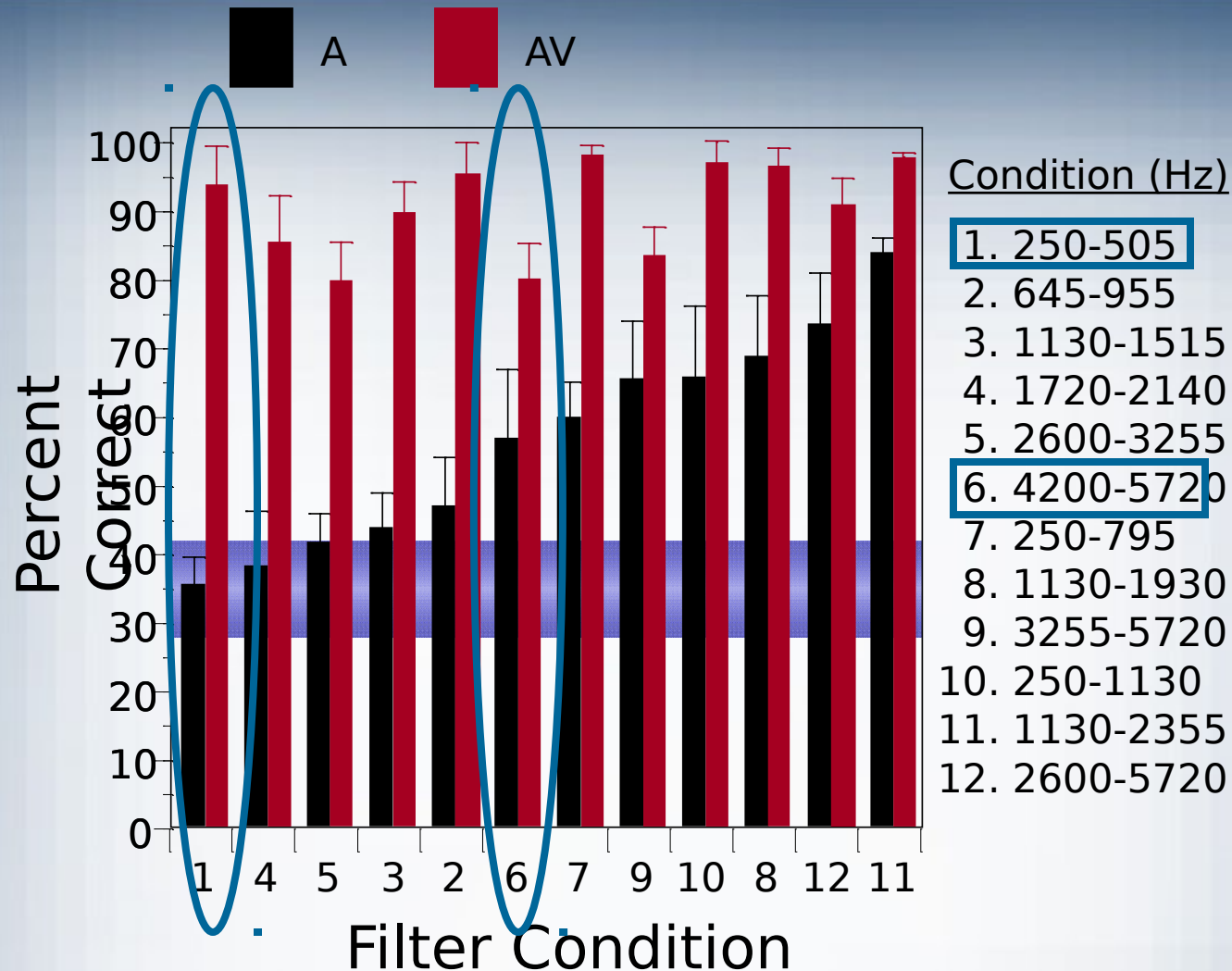
Correlation Method for Speech



Background (continued)

- Are the relative importance of different frequency regions altered by the presence of visual speech cues?
- Past results using isolated spectral bands of speech show that low-frequency speech provides more benefit to speechreading than other spectral regions (Grant and Walden, 1996).

Background (continued)



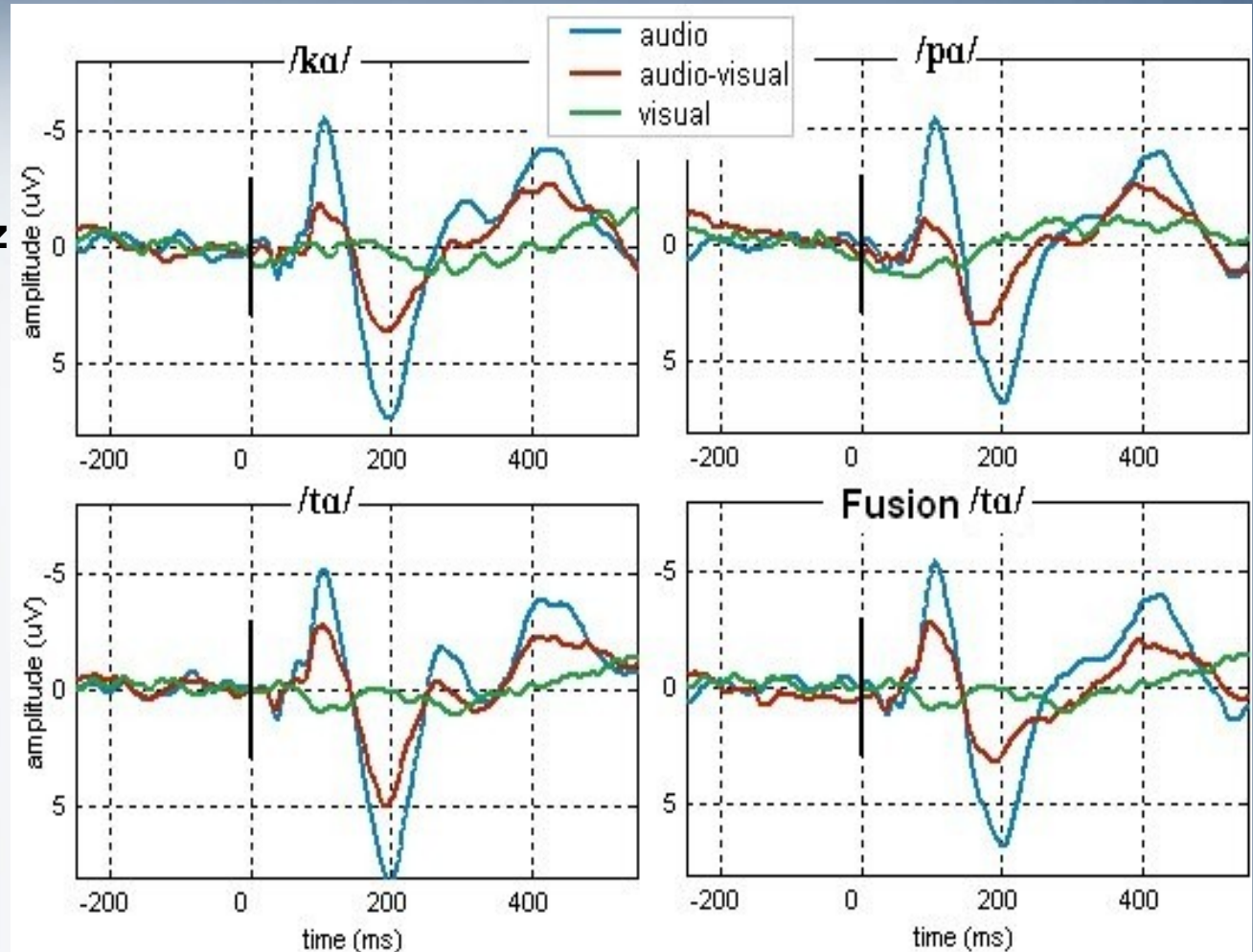
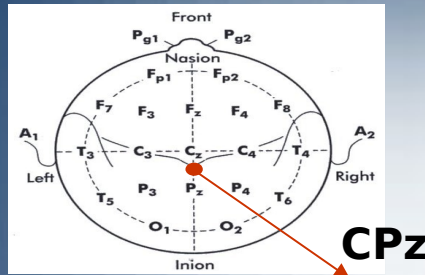
From Grant and Walden (1996). JASA, 100, 2415-

2424.

Background (continued)

- Evidence from electrophysiological studies show that visual speech cues fundamentally alter the way the auditory cortex responds to sound input (Calvert, 1977; van Wassenhove et al., 2005).
 - reduction in N1-P2 amplitude.
 - latency shift in N2 peak for highly visible consonants.

Visual Speech Alters Neural Processing of Auditory Speech

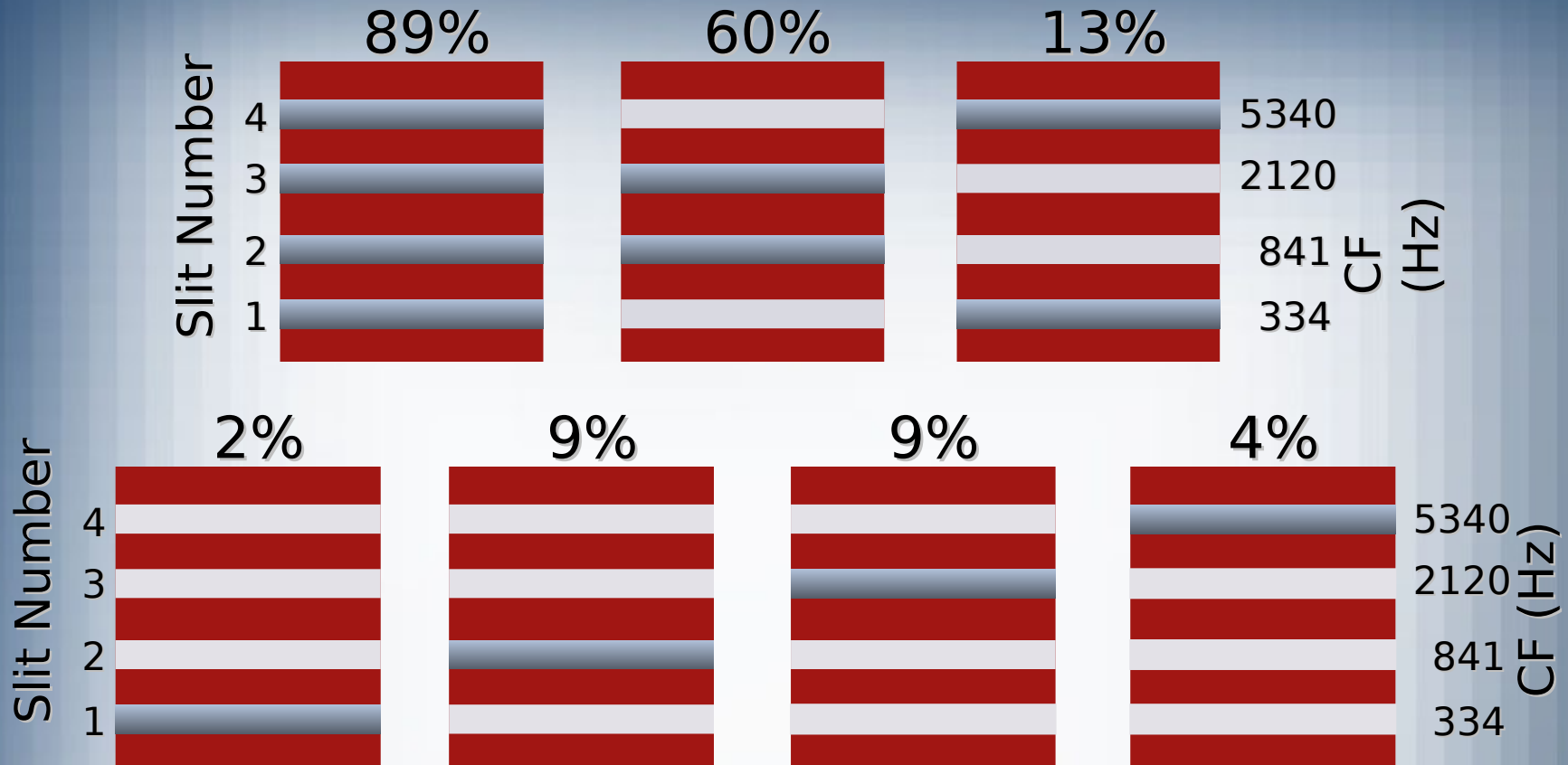


From van Wassenhove, Grant, and Poeppel (2005). PNAS, 102, 1181-

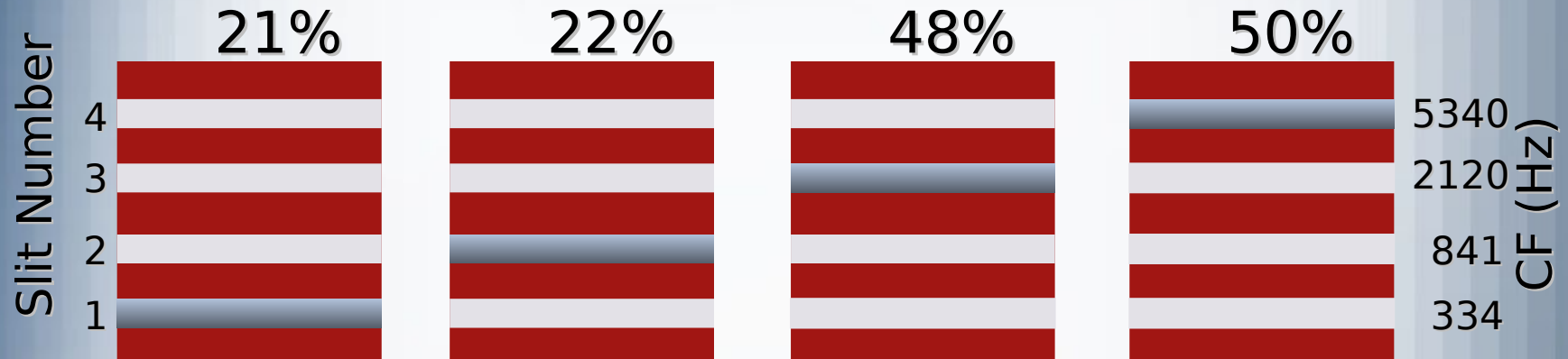
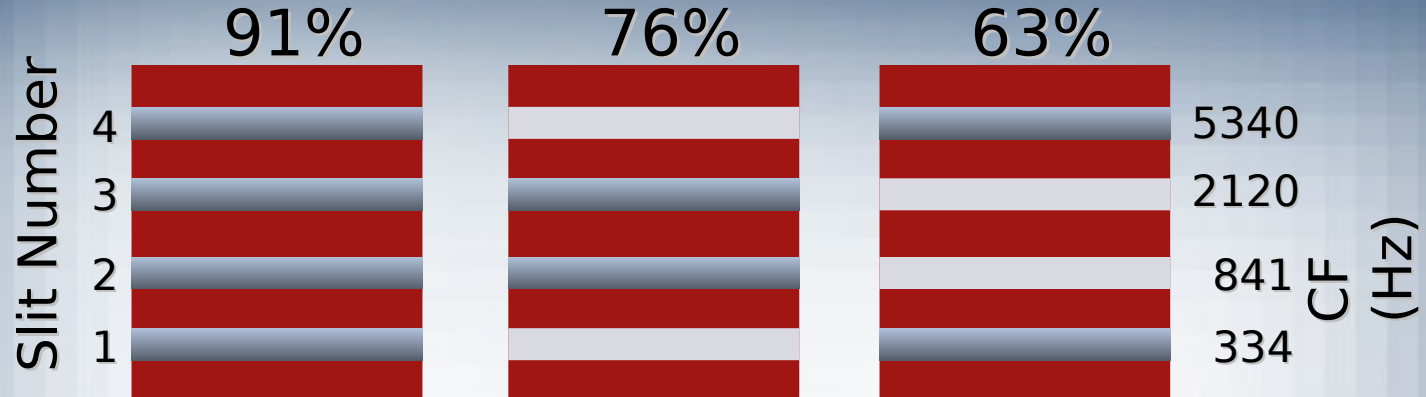
Goals

- Determine relative importance of different frequency regions for auditory and auditory-visual speech.
- Minimize band-on-band interactions by partitioning the speech signal into widely spaced narrow bands.

Spectral Slits - Sentences

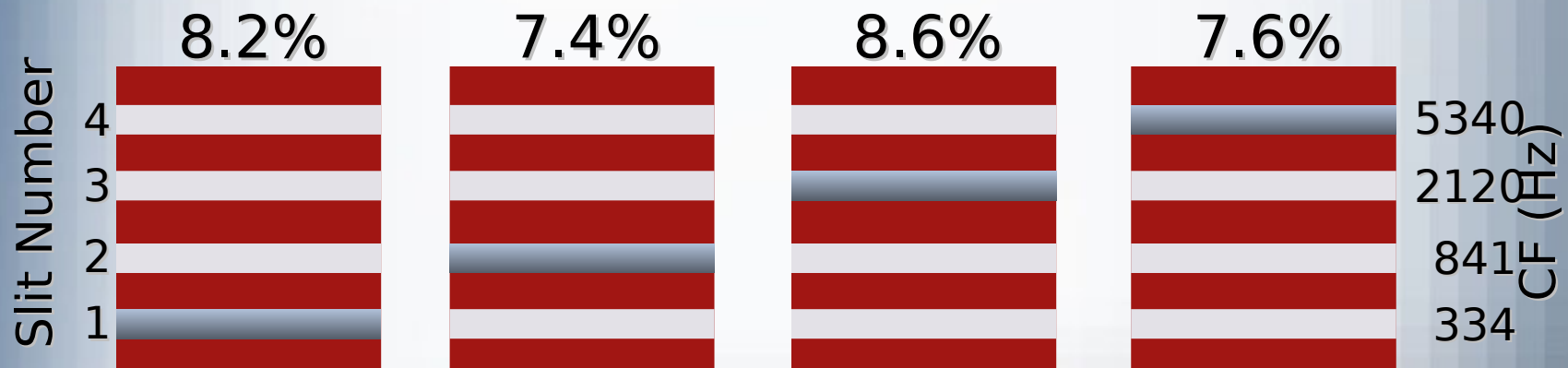


Spectral Slits - Consonants

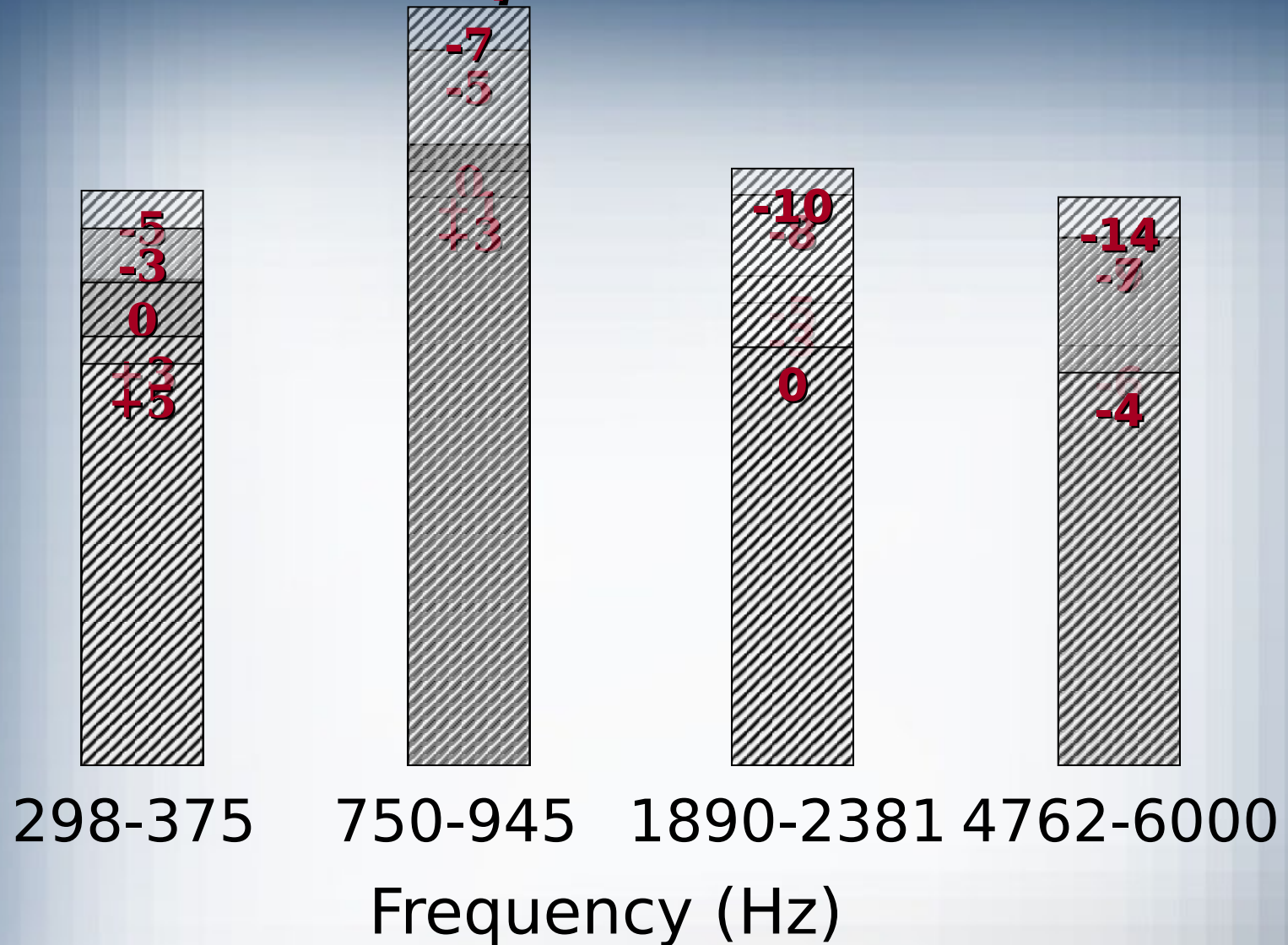


Spectral Slits - Consonants

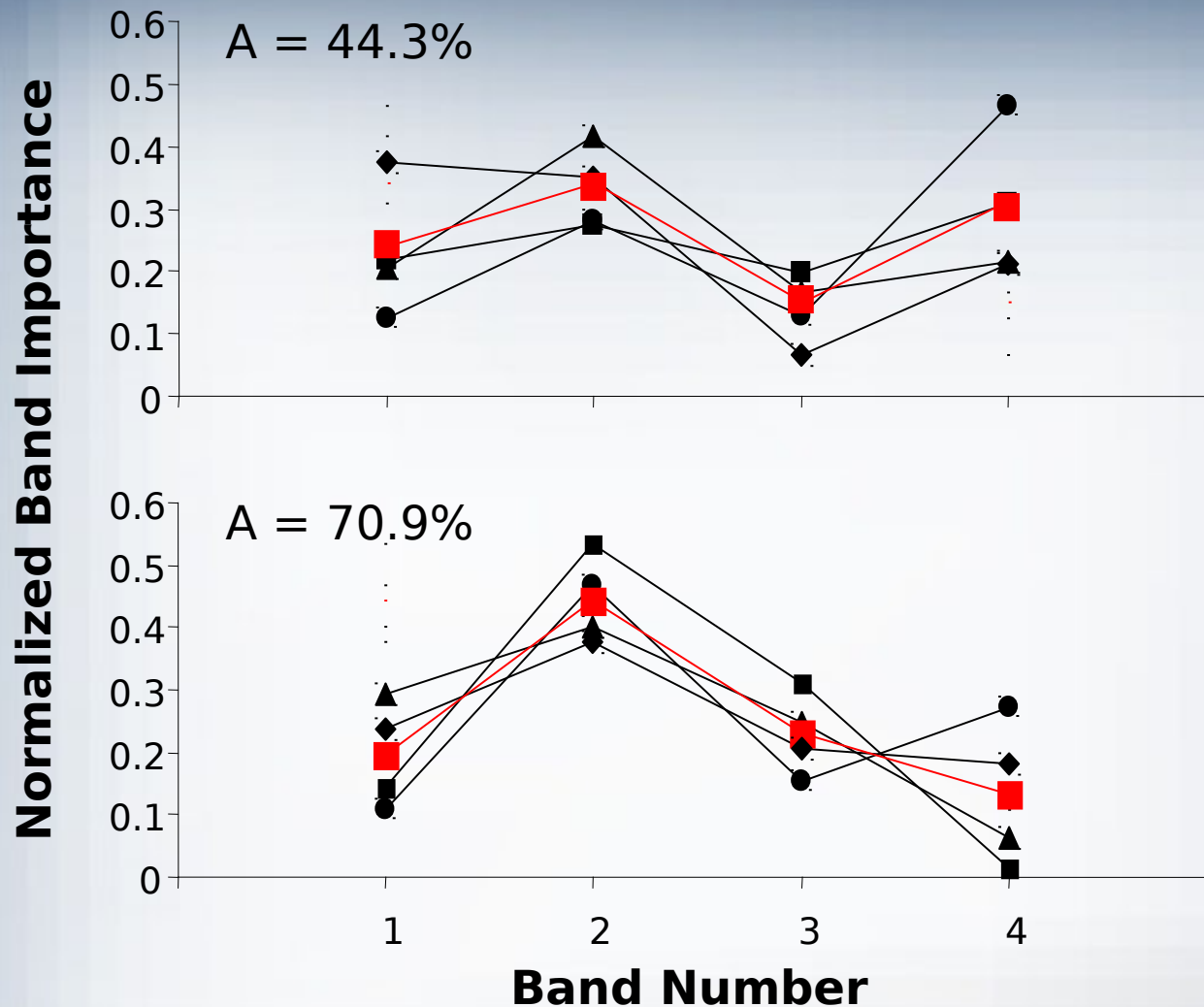
- Individual band scores are too high for AV testing. AV scores would be at ceiling.
- Different amounts of masking noise needed for each band.
- Goal in selecting noise levels was to:
 - make each band roughly equal in intelligibility.
 - make the the combination of all 4 bands roughly 40% intelligibile.



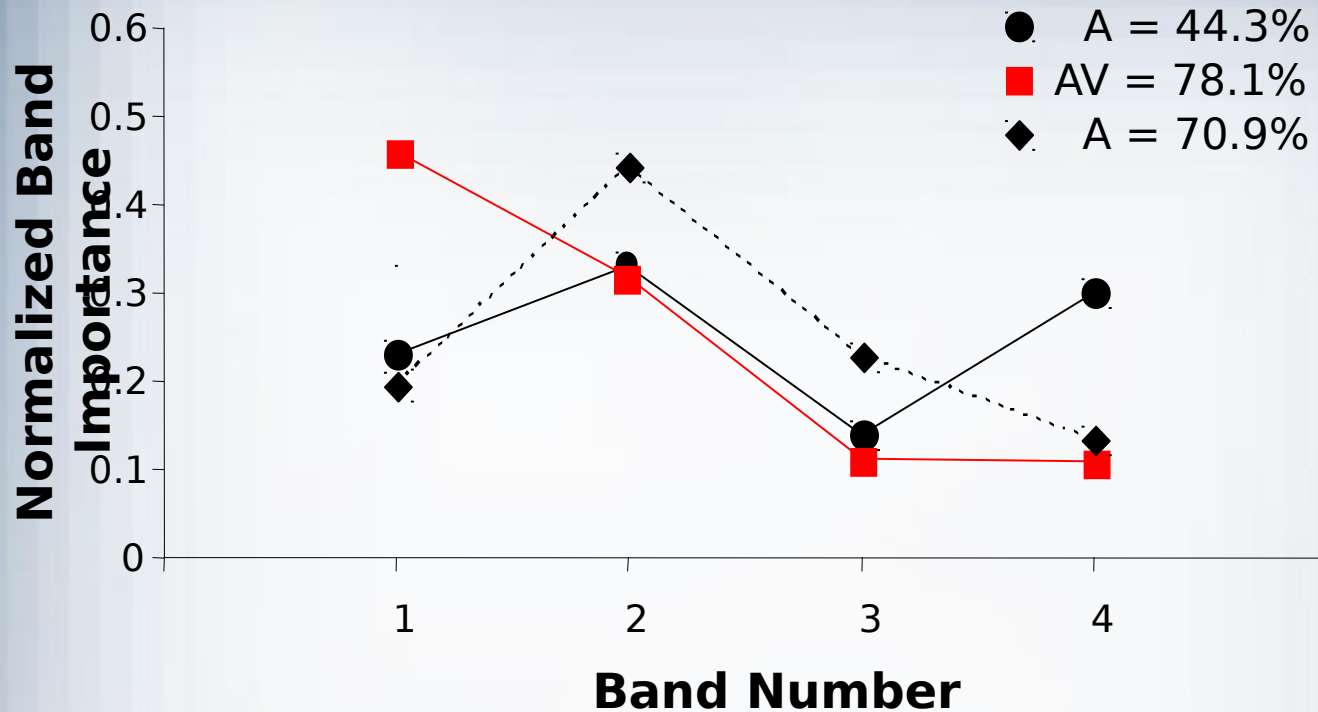
Correlation Method for Speech



Band Importance (Audio Alone)



Band Importance (A versus AV)



Discussion – Audio Alone

- Frequency-importance functions for auditory alone conditions show that listeners consistently weighted band 2 the greatest.
- Relative importance changed slightly when the overall intelligibility of the auditory condition was increased.
 - band 2 still given the greatest weight.
 - relative weight for bands 3 and 4 are swapped.

Discussion – Audiovisual

- When visual speech cues are present, listeners' place more importance on low frequencies.
- Results are consistent with past studies using isolated spectral bands of speech.
 - low-frequency speech provides cues for voicing which is highly complementary with speechreading.
 - mid-to-high-frequency speech provides cues for place of articulation which is highly redundant with speechreading.

Conclusions - Questions

- For robust speech recognition, information must be extracted from many different spectral regions.
- The presence or absence of visual speech cues alters the importance of different spectral regions for the listener.
- For listening conditions where low-frequency speech cues are compromised (noise, reverberation, hearing loss), enhancement of the low frequencies of speech may be advantageous, especially in situations where visual cues are available.